Chapter 7

Section 7.1

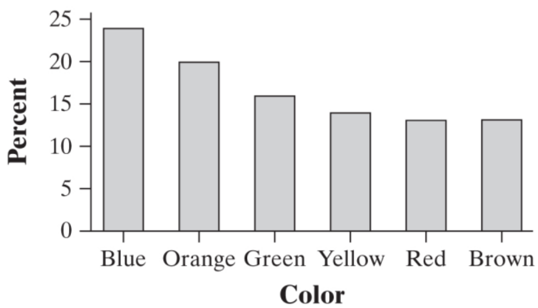
Check Your Understanding, page 425:

1. The parameter is $\mu = 20$ ounces of iced tea. The statistic is $\bar{x} = 19.6$ ounces of iced tea.

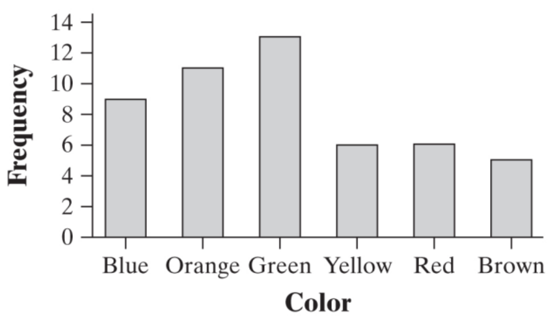
2. The parameter is $p = 0.10$, or 10% of passengers. The statistic is $\hat{p} = 0.08$, or 8% of the sample of passengers.

Check Your Understanding, page 428:

1. The individuals are the M&M'S Milk Chocolate Candies, the variable is the color of the M&M, and the parameter of interest is the proportion of orange M&M'S. The graph below shows the population distribution.



2. The graph below shows a possible distribution of sample data. For this sample there are 11 orange M&M'S so $\hat{p} = \frac{11}{50} = 0.22$.



3. The middle graph is the approximate sampling distribution of $\hat{p}$. The statistic measures the proportion of orange candies in samples of 50 M&M'S. Assuming that the company is correct, 20% of the M&M'S are orange, so the center of the distribution of $\hat{p}$ should be at approximately 0.20. The first graph shows the distribution of the colors for one sample, rather than the distribution of $\hat{p}$ from many samples, and the third graph is centered at 0.40 rather than 0.20.

Check Your Understanding, page 434:
1. The median does not appear to be an unbiased estimator of the population median. The mean of the approximate sampling distribution of the sample median (73.5) is not equal to the median of the population (75).
2. Increasing the sample size from 10 to 20 will decrease the spread of the sampling distribution. Larger samples provide more precise estimates because larger samples include more information about the population distribution.
3. The sampling distribution of the sample median is skewed to the left and unimodal.

Exercises, page 436:
7.1 (a) *Population*: all people who signed a card saying that they intend to quit smoking. *Parameter*: the proportion of the population who actually quit smoking. *Sample*: a random sample of 1000 people who signed the cards. *Statistic*: the proportion of the sample who actually quit smoking; $\hat{p} = 0.21$.
(b) *Population*: all the turkey meat. *Parameter*: minimum temperature in all of the turkey meat. *Sample*: four randomly chosen locations in the turkey. *Statistic*: minimum temperature in the sample of four locations; sample minimum = 170°F.

7.2 (a) *Population*: individuals in U.S. households. *Parameter*: proportion of the U. S. population who were unemployed. *Sample*: a random sample of individuals from 60,000 U.S. households. S*tatistic*: the proportion of the sample who were unemployed; $\hat{p}$ = 0.079.
(b) *Population*: all gasoline stations in a large city. *Parameter*: range of gas prices at all gasoline stations in the city. *Sample*: a random sample of 10 gas stations in the city. S*tatistic*: the range of prices in the sample; sample range = 25 cents.

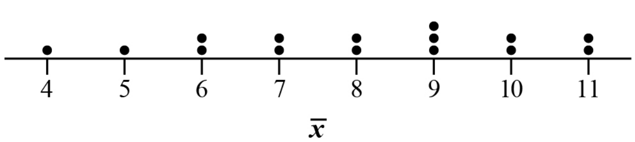
7.3 $\mu = 2.5003$ is a parameter (related to the population of all the ball bearings in the container) and $\bar{x} = 2.5009$ is a statistic (related to the sample of 100 ball bearings).

7.4 $p = 0.41$ is a parameter (related to the population of all registered voters) and $\hat{p} = 0.33$ is a statistic (related to the sample of 250 registered voters).

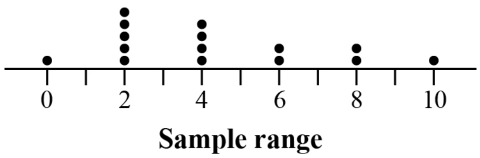
7.5 $\hat{p} = 0.48$ is a statistic (related to the sample of 100 numbers dialed) and $p = 0.52$ is a parameter (related to the population of all residential phone numbers in the city).

7.6 $\bar{x} = 64.5$ inches is a statistic (related to the sample of female college students) and $\mu = 63$ inches is a parameter (related to the population of all adult American women).

7.7 (a) The possible SRSs are 2 and 6 ($\bar{x}$ = 4), 2 and 8 ($\bar{x}$ = 5), 2 and 10 ($\bar{x}$ = 6), 2 and 10 ($\bar{x}$ = 6), 2 and 12 ($\bar{x}$ = 7), 6 and 8 ($\bar{x}$ = 7), 6 and 10 ($\bar{x}$ = 8), 6 and 10 ($\bar{x}$ = 8), 6 and 12 ($\bar{x}$ = 9), 8 and 10 ($\bar{x}$ = 9), 8 and 10 ($\bar{x}$ = 9), 8 and 12 ($\bar{x}$ = 10), 10 and 10 ($\bar{x}$ = 10), 10 and 12 ($\bar{x}$ = 11), 10 and 12 ($\bar{x}$ = 11).
(b) The sampling distribution of $\bar{x}$ is skewed to the left and unimodal. The mean of the sampling distribution is 8, which is equal to the mean of the population. The values of $\bar{x}$ vary from 4 to 11.



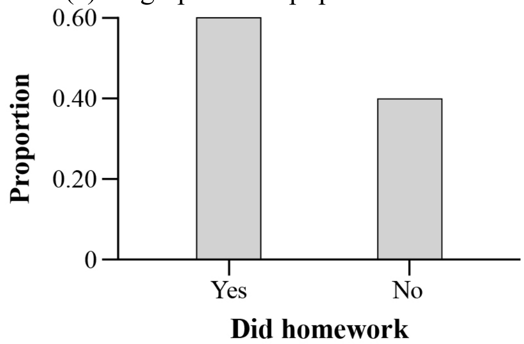
7.8 (a) The possible SRSs are 2 and 6 (range = 4), 2 and 8 (range = 6), 2 and 10 (range = 8), 2 and 10 (range = 8), 2 and 12 (range = 10), 6 and 8 (range = 2), 6 and 10 (range = 4), 6 and 10 (range = 4), 6 and 12 (range = 6), 8 and 10 (range = 2), 8 and 10 (range = 2), 8 and 12 (range = 4), 10 and 10 (range = 0), 10 and 12 (range = 2), 10 and 12 (range = 2).
(b) The sampling distribution of the sample range is skewed to the right and unimodal. The mean of the sampling distribution is 4.27, which is much less than the range of the population (10). The values of the sample range vary from 0 to 10.



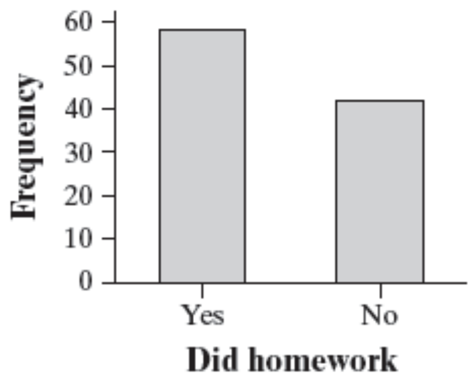
7.9 (a) In one simulated SRS of 100 students, there were 73 students who did all of their assigned homework last week.
(b) The distribution is reasonably symmetric and bell-shaped. It is centered at about 0.60. Values vary from about 0.47 to 0.74. There don't appear to be any outliers.
(c) Because there were no values of $\hat{p}$ less than or equal to 0.45 in the simulation, it would be very surprising to get a sample proportion of 0.45 or less in an SRS of size 100 from a population in which $p = 0.60$.
(d) Because it would be very surprising to get a sample proportion of 0.45 or less in an SRS of size 100 when $p = 0.60$, we should be skeptical of the newspaper's claim. Sampling variability is not a plausible explanation for the difference between the observed proportion and the proportion claimed by the newspaper.

7.10 (a) In one simulated SRS of 20 students, the average height was $\bar{x} = 62.4$ inches.
(b) The distribution is reasonably symmetric and bell-shaped. It is centered at about 64. Values vary from about 62.4 to 65.7. There do not seem to be any outliers.
(c) Because about 10% of the values of $\bar{x}$ were 64.7 or greater, it wouldn't be surprising to get a sample mean of 64.7 or larger in an SRS of size 20 from a Normal population with $\mu = 64$ and $\sigma = 2.5$.
(d) Because it isn't surprising to get a sample mean of 64.7 or greater in an SRS of size 20 from a Normal population with $\mu = 64$ and $\sigma = 2.5$, we do not have convincing evidence that the population mean height at the school is different than $\mu = 64$. Sampling variability is a plausible explanation for the difference between the sample mean and the mean claimed by the National Center for Health Statistics.

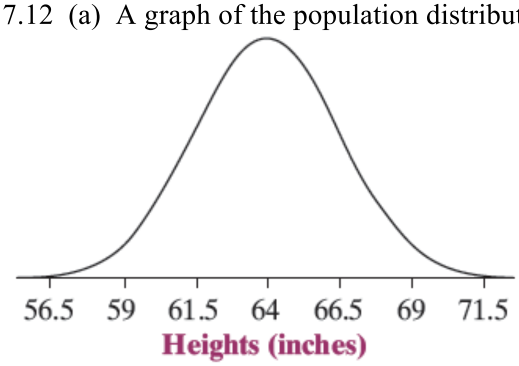
7.11 (a) A graph of the population distribution is shown below.



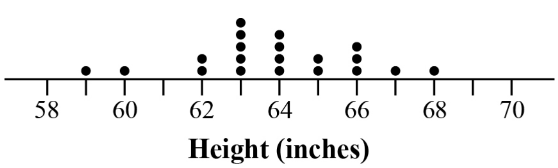
(b) Answers will vary. An example bar graph is given.



7.12 (a) A graph of the population distribution is shown below.



(b) Answers will vary. An example dotplot is given below.



7.13 (a) The approximate sampling distribution is skewed to the right with a center at $9\left({}^{\circ}\text{F}\right)^2$. The values vary from about 2 to 27.5 $\left({}^{\circ}\text{F}\right)^2$.

(b) A sample variance of 25 $\left({}^{\circ}\mathrm{F}\right)^2$ is quite large compared with what we would expect, because only one out of 500 SRSs had a variance that high. A sample variance of 25 provides convincing evidence that the manufacturer's claim is false and that the thermostat actually has more variability than claimed.

7.14 (a) The approximate sampling distribution is slightly left-skewed and centered at $45.5^{\circ}\mathrm{F}$. The values vary from about 39 to $50^{\circ}\mathrm{F}$.
(b) A sample minimum of $40^{\circ}\mathrm{F}$ is quite low compared with what we would expect, because very few of the 500 SRSs had a minimum that low. A sample minimum of 40ºF provides convincing evidence that the manufacturer's claim is false.

7.15 If we chose many SRSs and calculated the sample mean $\bar{x}$ for each sample, the distribution of $\bar{x}$ would be centered at the value of μ. In other words, when we use $\bar{x}$ to estimate μ, we will not consistently underestimate μ or consistently overestimate μ.

7.16 If we chose many random samples and calculated the sample proportion $\hat{p}$ for each sample, the distribution of $\hat{p}$ would be centered at the value of *p*. In other words, when we use $\hat{p}$ to estimate *p*, we will not consistently underestimate *p* or consistently overestimate *p*.

7.17 A larger random sample will provide more information and, therefore, more precise results. The variability of the distribution of $\bar{x}$ decreases as the sample size increases.

7.18 A larger random sample will provide more information and, therefore, more precise results. The variability of the distribution of $\hat{p}$ decreases as the sample size increases.

7.19 (a) Statistics ii and iii are unbiased estimators because the means of their sampling distributions appear to be equal to the corresponding population parameters.
(b) Statistic ii does the best job at estimating the parameter. It is unbiased and has very little variability

7.20 (a) It would be better to use an SRS of 20,000 returns for estimating the proportion of all returns claiming itemized deductions. Larger samples provide more precise estimates because larger samples include more information about the population distribution.

7.21 c.

7.22 d.

7.23 a.

7.24 b.

7.25 (a) We are looking for the percentage of values that are 2.5 standard deviations or farther below the mean in a Normal distribution. In other words, we are looking for $P(Z \le -2.5)$. Using Table A, $P(Z \le -2.5) = 0.0062$. *Using technology*: normalcdf(lower: –1000, upper: –2.50, μ : 0, σ : 1) = 0.0062. Less than 1% of healthy young adults have osteoporosis.
(b) Let X be the BMD for women aged 70 to 79 on the standard scale. Then X follows a $N(-2, 1)$ distribution and we want to find $P(X \le -2.5)$. The standardized score for the boundary value is

$z = \dfrac{-2.5-(-2)}{1} = -0.5$ and the probability is $P(Z \le -0.5) = 0.3085$. Using technology: normalcdf(lower: –1000, upper: –2.5, μ : –2, σ = 1) = 0.3085. About 31% of women aged 70–79 have osteoporosis.

7.26 (a) The equation for the least squares regression line is: $\hat{y} = 1.4146 + 0.4399x$ where $\hat{y}$ is the predicted average number of offspring per female and x is the index of the abundance of pine cones.
(b) The linear model is appropriate for these data because there is no leftover pattern in the residual plot.
(c) r^2 : 57.2% of the variability in the average number of offspring can be accounted for by the linear model relating average number of offspring to abundance of pine cones. *s*: When using this model to predict average number of offspring from abundance of pine cones, we will typically be off by about 0.60 offspring.